

ПРОГНОЗИРОВАНИЕ ИНДЕКСА ПОТРЕБИТЕЛЬСКИХ ЦЕН ТОМСКОЙ ОБЛАСТИ С ИСПОЛЬЗОВАНИЕМ МОДЕЛИ ARIMA

*Е.А. Кочегурова, к.т.н., доц. ОИТ ИШИТР,
Я.А. Михайлова, студент, гр. 8И8Б
Томский политехнический университет
E-mail: kocheg@tpu.ru*

Введение

Индекс потребительских цен (ИПЦ) является одним из важнейших экономических показателей, характеризующих инфляцию в стране. ИПЦ показывает изменение цен на товары и услуги с точки зрения потребителей. В современных макроэкономических условиях прогнозирование уровня ИПЦ является важной практической задачей. В качестве модели временного ряда (ВР) для предсказания уровня ИПЦ в данной работе выбрана ARIMA-модель. Эта модель является результатом комбинации авторегрессии $AR(p)$, интегрирования и скользящего среднего $MA(q)$ и зависит от параметров: p – порядка авторегрессии, d – порядка интегрирования, q – порядок скользящего среднего.

Подготовка данных

Исходный ВР представляет собой ИПЦ на товары и услуги в Томской области за период (01.2014 – 12.2019). В данном исследовании использована облачная платформа Google Colaboratory и язык программирования – Python. Помимо часто используемых математических библиотек Python, таких как NumPy, используются также специализированные библиотеки. Scikit-learn – один из наиболее широко используемых пакетов Python для машинного обучения и науки о данных. Statsmodels – это пакет Python, который позволяет пользователям исследовать данные, оценивать статистические модели и выполнять статистические тесты.

ВР состоит из 72 значений, из которых 10% (7 последних значений) используется в качестве выборки для прогноза. Соответственно, 65 первых значений ВР предназначены для обучения модели. График тренировочной части временного ряда представлен на рисунке 1.

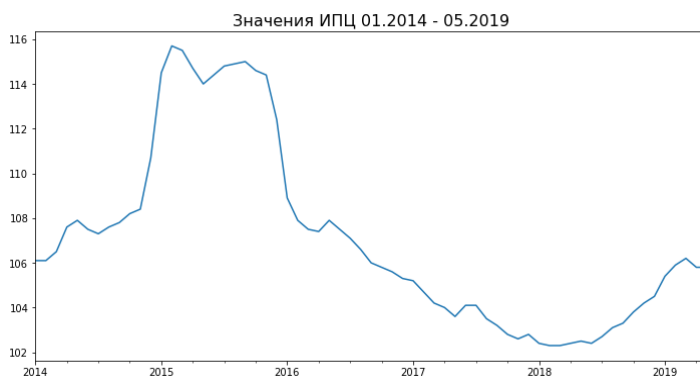


Рис. 1. График временного ряда ИПЦ

Подбор модели и прогноз

Согласно концепции алгоритма ARIMA, первым этапом является проверка нулевой гипотезы о нестационарности ряда. И в соответствии со статистическим тестом Дики-Фуллера в выборке ИПЦ присутствует тренд: тестовое значение больше критического ($-1,49 > -2,597$) при уровне значимости 10%. Также p -значение больше 0.05, что тоже можно интерпретировать как отказ от отвержения нулевой гипотезы. После выделения тренда тест Дики-Фуллера, проведенный над приращениями ВР, позволяет считать ряд стационарным. Таким образом, параметр интегрирования d для модели ARIMA должен быть по крайней мере 1.

Для подбора параметров авторегрессии и скользящего среднего проанализированы графики автокорреляционной функции (АКФ, Autocorrelation Function) и функции частичной автокорреляции (ЧАКФ, Partial Autocorrelation Function), представленные на рисунке 2.

На основании графиков АКФ и ЧАКФ определены параметры скользящего среднего и авторегрессии, равные $p=2$ и $q=1$. Так получили модель для прогноза $ARIMA(2, 1, 1)$.

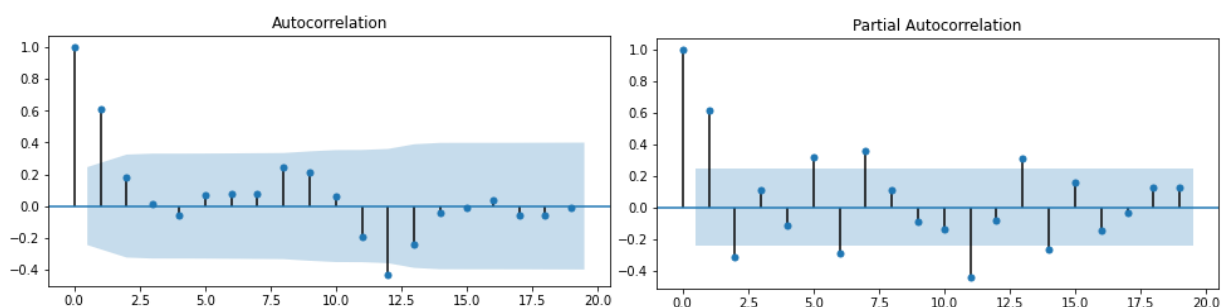


Рис. 2. АКФ и ЧАКФ

Кроме стандартной методики выбора параметров ARIMA-модели была выполнена настройка параметров модели на основе сканирования в области допустимых значений. В результате – были найдены такие комбинации параметров p , d и q , которые не приводят к ошибке моделирования, и из них отбирается лучшая. Оцениваются модели на основании RMSPE-погрешности (Root Mean Square Percentage Error). Используемый в данной работе алгоритм делит BP на тренировочную и тестовую часть в соотношении 50:50. Для каждого значения тестовой выборки создаётся модель на основании поданных на вход модели параметров p , d и q и прошлых значений ряда. Далее выполняется краткосрочный прогноз, при этом тренировочная выборка обновляется. На основании такого поиска на сетке наилучшей является модель ARIMA(1, 1, 0). Данная модель показала лучшие результаты по сравнению с моделью, параметры которой подобраны на основании стандартной методики. Ниже представлен график с результатом прогноза на основе модели ARIMA(1, 1, 0).

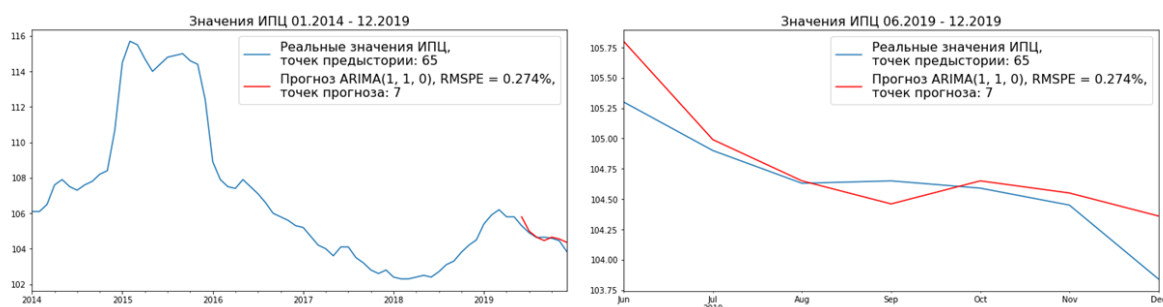


Рис. 3. Графики прогноза ИПЦ

RMSPE-погрешность данной модели для BP ИПЦ составляет 0,274%, что можно считать приемлемым результатом. На графике видно, что спрогнозированные значения не идеально близки к реальным, однако так же хорошо видно, что прогнозный график повторяет тенденцию графика реальных значений.

Заключение

В результате работы был реализован алгоритм прогнозирования ИПЦ по Томской области на товары и услуги на основе ARIMA-модели. Хочется отметить, что задача прогнозирования временных рядов может быть сопряжена со сложностью интерпретации характеристик ряда, выбора метода и подбора параметров моделей.

Список использованных источников

1. Кочегурова, Е.А. Гибридный подход для краткосрочного прогнозирования временных рядов на основе штрафного Р-сплайна и эволюционной оптимизации /Е.А. Кочегурова, Е.Ю. Репина, О.Б. Цехан // Компьютерная оптика. – 2020. – Т. 44, № 5. – С. 821- 829. – DOI: 10.18287/2412-6179-CO-667.
2. Сапова А.К. Прогнозирование инфляции на основе индекса потребительских цен с учетом влияния сезонного фактора // Статистика и экономика. 2017. №6. URL: <https://cyberleninka.ru/article/n/prognostirovanie-inflyatsii-na-osnove-indeksa-potrebitelskih-tsen-s-uchetom-vliyaniya-sezonnogo-faktora> (дата обращения: 02.03.2021).
3. Чучуева И.А. МОДЕЛЬ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ ПО ВЫБОРКЕ МАКСИМАЛЬНОГО ПОДОБИЯ, диссертация... канд. тех. наук / Московский государственный технический университет им. Н.Э. Баумана. Москва, 2012.